



TAMPERE ECONOMIC WORKING PAPERS  
NET SERIES

TOP INCOMES AND TOP TAX RATES: IMPLICATIONS  
FOR OPTIMAL TAXATION OF TOP INCOMES IN FINLAND

Marja Riihelä  
Risto Sullström  
Matti Tuomala

Working Paper 88  
May 2014

SCHOOL OF MANAGEMENT  
FI-33014 UNIVERSITY OF TAMPERE, FINLAND

ISSN 1458-1191  
ISBN 978-951-44-9402-4

# Top Incomes and Top Tax Rates: Implications for Optimal Taxation of Top Incomes in Finland

Marja Riihelä - Risto Sullström - Matti Tuomala

## Abstract

We apply optimal top marginal tax rate formulas with alternative social preferences to the Finnish economy using evidence on the responsiveness of top incomes in Finland to changes in top tax rates that have taken place in Finland over the last 30 years. Based on the Finnish income distribution data (cross section) we estimated by using maximum likelihood method several two and three parameter distributions. Among two parameter distributions the Champnowne one is the best fitting for the pre-tax income distribution in Finland (1990-2010). We also recognize that there is much uncertainty particularly related to labour supply elasticity and do not simply rely on the central point estimate. It is safe to conclude from our application that the current top marginal tax rate in Finland is not close to the top of the Dupuit-Laffer curve.

*Key words:* top tax rates; optimal taxation; individual incomes

*JEL classification:* C63; H21; H24

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Optimal top income tax rates in the Mirrlees model</b>	<b>3</b>
2.1	Utilitarian social preferences and top tax rates . . . . .	6
2.2	The top tax rate and income effects . . . . .	7
2.3	The asymptotic solution . . . . .	7
<b>3</b>	<b>The taxable income elasticity at the top in Finland</b>	<b>8</b>
<b>4</b>	<b>Distribution of top incomes: Estimation of the Pareto and Champernowne-Fisk distributions</b>	<b>14</b>
<b>5</b>	<b>Optimal top rates for the different assumptions in Finland</b>	<b>17</b>
<b>6</b>	<b>Concluding remarks</b>	<b>21</b>
	<b>References</b>	<b>22</b>
	<b>Appendix A</b>	<b>27</b>
	<b>Appendix B</b>	<b>28</b>
	<b>Appendix C</b>	<b>31</b>
	<b>Appendix D</b>	<b>34</b>

# 1 Introduction

The increasing share of the top income earners in total income has been a notable feature of the changes in income inequality in English speaking countries while in Europe Netherlands, France and Switzerland and Japan display hardly any change in top income shares<sup>1</sup>. This trend toward income concentration has also taken place in the Nordic countries, traditionally low inequality countries. The top percentile disposable income share in Finland doubled in the latter part of 1990s. At the same time, top tax rates on upper income earners have declined significantly in many OECD countries, again particularly in English speaking countries.<sup>2</sup> This is also the case in Finland (see Figures 1 and 2). Economists have formulated several hypotheses about causes of increasing inequality, but there is not a fully compelling explanation. For example, Atkinson et al. (2011) emphasise that it's very difficult to account for these figures with the standard labour supply, labour demand explanation. Hence we really have to think about things like social policies and progressive taxation.

Optimal income tax literature provides a striking result on the top marginal tax rate. The optimal marginal tax rate for the highest-wage person is zero. This result - due to Phelps-Sadka - really says that the highest income that could possibly happen should be subject to a zero marginal tax rate. Strictly speaking, this result applies only to a single person at the very top of the income distribution, suggesting it is a mere theoretical curiosity. Moreover, it is unclear that a "top earner" even exists. For example, Saez (2001) argues that "unbounded distributions are of much more interest than bounded distributions to address the high income optimal tax rate problem." Without a top earner, the intuition for the zero top marginal rate does not apply, and marginal rates near the top of the income distribution may be positive and even large. Calculations in Tuomala (1984) show that the zero rate is not a good approximation for high incomes.

In Mirrlees (1971) all wage distributions were unbounded above and therefore he did not have a zero rate result. He in turn presented precise conjectures about optimal tax rates in the case of utility functions separable in consumption and labour. One of the least well known features of Mirrlees (1971) paper is the demonstration (pages 189-200) that the optimal marginal tax rate converges to a positive value when the upper tail of the skill dis-

---

<sup>1</sup>The more recent estimates of Camille Landais (2007) show a rise in recent years in France.

<sup>2</sup>Piketty et al. (2011) investigates the link between skyrocketing inequality and top tax rates in OECD countries. They find a strong correlation between tax cuts for the highest earners and increases in the income share of the top 1 per cent since 1975.

tribution is of Paretian form, with this value being a function of the Pareto parameter and the characteristics of the utility function. It appears that the role of the latter depends solely on their appearance in the constraints, and does not depend on them entering the government's objective function. It is also important to note that the key assumptions behind asymptotic marginal tax results are that either the marginal utility of consumption or the social marginal valuation of utility goes to zero when the wage rates tend to infinity (see also Dahan and Strawczynski, 2012). In this situation we need only information on labour supply elasticities and the shape of the skill distribution to determine the optimal top marginal income tax rate.

The distribution of top incomes is a central part in determining top marginal tax rates. Tony Atkinson (2012, p. 774) wrote recently:

*“Economists tend to assume that it is  $e$  (the elasticity) that is the core of their subject, but equally central should be  $a$  (the distribution). This is particularly the case where the distribution of top incomes is becoming more concentrated in the form of a lower value for  $a$  (the Pareto parameter), implying a higher optimal top tax rate”.*

Because one of the key factors explaining the shape of an optimal income tax schedule and top marginal tax rates is the assumed family of distributions of earning abilities, it is of interest to look at other distributions than the lognormal and the Pareto distribution. As commonly known the lognormal distribution fits reasonable well over a large part of income range but diverges markedly at both tails. The Pareto distribution in turn fits well at the upper tail. Champernowne (1952) proposes a model in which individual incomes were assumed to follow a random walk in the logarithmic scale. Tuomala (2006) replaces the lognormal distribution by the Champernowne distribution<sup>3</sup>. Specifically he uses the two parameter version of the Champernowne distribution, also known as the Fisk-distribution. This distribution approaches asymptotically a form of Pareto distribution for large values of wages but it also has an interior maximum.

Finally, the relevant elasticities are crucial for optimal marginal tax rates. If high-income workers are particularly elastic in how their taxable income decreases with higher tax rates, this would imply lower optimal marginal tax rates on high incomes, all else the same. Saez (2001) used a constant labour supply elasticity formulation not because there was strong empirical evidence for it, but because there was no strong evidence against it. However, since the Saez (2001) paper, the survey by Røed and Ström (2002) provides some evidence for labour supply elasticity declining with income for Nor-

---

<sup>3</sup>It is often - also in an earlier version of this paper - referred to as the Fisk distribution (see Fisk, 1961 and also Bevan, 2005).

way. Røed and Ström (2002) offer a review of the evidence. They conclude that the limited evidence indicates that labour supply elasticities are declining with household income. Using Norwegian data Aaberge and Colombino (2006) provides support for declining elasticities. By contrast, there is empirical evidence on the elasticity of taxable income that higher elasticities are among high income individuals. Feldstein (1995) estimated large elasticities of taxable income with respect to tax rates among high earners. Gruber and Saez (2002) subsequently estimated smaller elasticities, but their estimates also support the hypothesis that the elasticity increases with income. But as with the distribution of abilities and the social welfare function, there is much debate over the true pattern of elasticities by income.

In this paper we apply top (asymptotic) marginal tax rate formulas to the Finnish case using empirical estimates on taxable income elasticity estimates and the distribution of pre-tax income. In the empirical part of this paper we derive the optimal marginal tax rates using evidence on the responsiveness of top incomes in Finland to changes in tax rates, based on the response of top incomes that have taken place in Finland over the last 30 years. We focus on two key tax reforms of 1988/89 and 1993.

The remainder of the paper is organized as follows. Section 2 analyzes optimal marginal top tax rates in the Mirrlees model in the case of quasi-linear preferences, a Pareto distribution of skills and constant labour supply elasticities. We also analyze an asymptotic solution in the case with the Cobb-Douglas preferences and the distribution with the Pareto tail. Section 3 describes the data and estimation results on labour supply elasticities and pre-tax inequality. In Section 4 we estimate the distribution of top incomes for the Finnish case. Section 5 applies the optimal top tax formulas in order to assess what the optimal taxation of top incomes is in Finland. Section 6 concludes.

## 2 Optimal top income tax rates in the Mirrlees model

There are a continuum of taxpayers, each having the same preference ordering, which is represented by a utility function

$$u = U(x) + V(1 - y) = x + V(1 - y) \quad (1)$$

where  $x$  is a composite consumption good and hours worked are  $y$ , with  $U_x = 1$  and  $V_y < 0$  (subscripts indicating partial derivatives) and where  $V(\cdot)$  is convex. Workers differ only in their exogenously given productivity or ability, denoted by  $n$  they can earn. There is a distribution of  $n$  on

the interval  $[0, \infty)$  represented by the density function  $f(n)$ . Gross income  $z = ny$ .

Suppose that the aim of policy can be expressed as maximizing the following social welfare criterion

$$W = \int_0^\infty G(u(n)) f(n) dn, \quad (2)$$

where  $G(\cdot)$  is an increasing and concave function of utility. The government cannot observe individuals' productivities and thus is restricted to setting taxes and transfers as a function only of earnings,  $T(z(n))$ .  $x = ny - T(ny)$  where  $T(\cdot)$  symbolizes the income tax, which is defined on total income since the wage  $n$  and the supplied amount of labour  $y(n)$  are not observed by the government.

The government maximizes  $W$  subject to the revenue constraint

$$\int_0^\infty T(z(n)) f(n) dn = R, \quad (3)$$

where in the Mirrlees tradition  $R$  is interpreted as the required revenue for essential public goods. The more non-tax revenue a government receives from external sources, the lower is  $R$ . In addition to the revenue constraint, the government faces incentive compatibility constraints. These in turn state that each  $n$  individual maximizes utility by choice of hours of work.

Diamond (1998) shows that when preferences satisfy (1) and labour supply elasticity  $\epsilon$  is constant<sup>4</sup>, the optimal marginal taxes must satisfy (see the derivation in the appendix A)

$$\frac{t}{1-t} = \left[ 1 + \frac{yV_{yy}}{V_y} \right] \left[ \frac{[1 - F(n)]}{nf(n)} \right] \left[ \frac{\int_n^\infty [1 - \varphi] f(p) dp}{(1 - F(n))} \right]. \quad (4)$$

Denoting by  $\lambda$  the multiplier of the government budget constraint we define the social marginal welfare weight on taxpayer  $n$  as  $\varphi_n = \frac{G'(u)}{\lambda}$ , where  $\lambda = \int_0^\infty G'[U(x)] f(p) dp$ . In other words the Lagrange multiplier  $\lambda$  is equal to the population average of  $G'[u(n)]$ . Hence welfare weights  $\varphi_n = \frac{G'(u)}{\lambda}$  are average to one.

For any social welfare function  $G$  with a property that  $\lim_{u \rightarrow \infty} G'(u) = 0$ , and individual preferences represented by  $u = x - [y^{1+\frac{1}{\epsilon}}]/[1 + \frac{1}{\epsilon}]$ , then the integral in (4) asymptotically converges to 1 and  $\frac{t(n)}{1-t(n)}$  converges to

$$\frac{t(n)}{1-t(n)} = \left[ 1 + \frac{1}{\epsilon} \right] \frac{1 - F(n)}{nf(n)}.$$

---

<sup>4</sup>Note in the quasi-linear case  $\epsilon = 1 + 1/\epsilon^c$ , where  $\epsilon^c$  is the compensated elasticity of labour supply.

In the case of the unbounded Pareto distribution,  $f(n) = \frac{1}{n^{1+a}}$  for  $a > 0$ ,  $\frac{1-F(n)}{nf(n)} = \frac{1}{a}$  is constant. Assuming a Pareto distribution of skills above the modal skill (see Diamond, 1998), the asymptotic optimal marginal tax rate is given by

$$\lim_{n \rightarrow \infty} \frac{t(n)}{1-t(n)} = \left[1 + \frac{1}{\epsilon}\right] \frac{1}{a} \quad (5)$$

where  $a$  is the Pareto coefficient. Hence

$$t = \frac{1}{1 + a/\zeta} \quad \text{where } \zeta = \left[1 + \frac{1}{\epsilon}\right] \quad (6)$$

is an explicit formula for the optimal top income tax rate. Assuming that the average social marginal welfare weight among top bracket income earners is zero allows us to obtain an upper bound on the optimal top tax rate. Hence the formula (6) gives revenue maximizing tax rate. If we assume the “Rawlsian” social objective<sup>5</sup>, a Pareto distribution and preferences are quasi-linear then the pattern of marginal tax rates depends only on  $\frac{[1-F(n)]}{nf(n)}$ , that is, on the shape of the  $n$ -distribution. We have the same formula as in (6).<sup>6 7</sup>

Saez (2001) provides a simple derivation of top income tax rate  $t$  in terms of gross income  $z$  based on approximation. The government chooses  $t$  to maximize tax revenue  $R$  from the top bracket (as the government puts no marginal social welfare weight on top bracket earners)<sup>8</sup>;  $R = t(z_m(1-t) - z^*)$ .

---

<sup>5</sup>Maximizing utility of the worst off person in the society is not the original version of Rawls (1971). It is a kind of welfarist version of Rawlsian. “To interpret the difference principle as the principle of maximin utility (the principle to maximize the well-being of the least advantaged person) is a serious misunderstanding from a philosophical standpoint.” Rawls (1982).

<sup>6</sup>Hence using the Rawlsian social welfare function we do not obtain the rising part of the U-shaped marginal tax rates as in Diamond (1998).

<sup>7</sup>Here we have a version of the Rawlsian social welfare function, where the maximand is the welfare of the worst off individual. This is a very standard interpretation in public economics. We could, however, adopt a different version. Rawls (1971, p. 98) writes “One possibility is to choose a particular social position, say that of the unskilled worker, and then to count the least advantaged of those with the average income and wealth of this group, or less. The expectation of the lowest representative man is defined as the average taken over this whole class. Another alternative is a definition solely in terms of relative income and wealth with no reference to social position. Thus all persons with less than half of the median income and wealth may be taken as the least advantaged segment”. It is some interest to note that if the Pareto distribution applies over the whole range of  $n$  and the maximand of the Rawlsian social welfare function is the welfare of those below the poverty line, the optimal marginal tax rate is increasing up to the poverty line. The importance of this observation undermines the fact that the fit of Pareto distribution over the whole range of income turns out to be quite poor.

<sup>8</sup>Alternatively, assume one wants to increase the marginal tax rate from  $t$  to  $t + dt$  over



$z^*$  is a threshold and  $z_m$  is the mean income of those with incomes above  $z^*$ . In other words, if  $z^*$  is a threshold above which a hypothetical top rate of tax applies, then the mean income of those affected is  $z_m$ , and  $z_m - z^*$  is the amount of income over which the new tax rate applies.

The first order condition is  $t \frac{dz_m}{d(1-t)} = z_m - z^*$ . Dividing by  $z_m$  we have

$$\frac{tdz_m(1-t)}{(1-t)z_md(1-t)} = 1 - \frac{z^*}{z_m} \quad (7)$$

where at the left hand side we have the elasticity of income in the top bracket with respect to the net-of-tax rate  $1-t$ . This is essentially the same formula as the formula (14) in Tuomala (1985) in the linear optimal income tax model. The revenue maximizing top rate that applies above the threshold  $z^*$  is given by  $\frac{t}{1-t}\epsilon = 1 - \frac{z^*}{z_m}$  or

$$t = \frac{1}{1 + \epsilon/I} \quad (8)$$

where  $I = 1 - \frac{z^*}{z_m}$  is a measure of inequality. But if the distribution of top incomes has a Pareto distribution then  $t$  is independent of the threshold  $z^*$ . This will hence approximately be true for large  $z$ . Assuming that the average social marginal welfare weight among top bracket income earners is zero allows us to obtain an upper bound on the optimal top tax rate. Hence the formula (7) gives revenue maximizing tax rate. The Pareto distribution has an important connection with van der Wijk's law. The average income  $z_m$  above the level  $z^*$  is proportional to  $z^*$  itself, i.e. or

$$\frac{z^*}{z_m} = \frac{a-1}{a}$$

(see for example the discussion in Cowell, 1977). Hence we have  $t = \frac{1}{1+\epsilon a}$ .

## 2.1 Utilitarian social preferences and top tax rates

One may argue that it is too extreme to put zero asymptotic welfare weight at the top earners. For example Feldstein (2012) argues that it is repugnant to

---

some income bracket  $[z; z + dz]$ . Then tax revenues go from  $R$  to  $R + dR$ , with:

$$dR = [1 - H(z)] dt dz - \frac{h(z) dz t' \epsilon z dt}{1-t}$$

( $h(z)$  is the density function for labour income, and  $H(z)$  is the distribution function).  $dR = 0$  if and only if

$$\frac{t}{1-t} = \frac{1 - H(z)}{\epsilon z h(z)}.$$

set weights in this way. Alternatively we might assume that  $\varphi$  has a positive lower bound which is approached as  $n$  rises without limit. Of course it is not obvious how to determine a positive lower bound. Given the lower point of  $\varphi$  the optimal rate is now

$$t = (1 - \varphi)/(1 - \varphi + \epsilon a), \quad (9)$$

with the Pareto distribution. The optimal top marginal tax rate is decreasing with the elasticity  $\epsilon$  and the social marginal welfare weight on top earner  $\varphi$  and decreasing in the Pareto parameter  $a$ . This rate is also less than the revenue maximizing rate with the same inequality parameter and elasticity.

## 2.2 The top tax rate and income effects

Saez (2001) also shows that the main results of Diamond (1998) can be generalized to preferences with income effect. Saez (2001) presents a formula for the optimal top tax rate for a more general setting based on an approximation. The welfare-maximizing top rate in the highest tax bracket equals  $t = (1 - \varphi)/(1 - \varphi a \epsilon^c - \eta)$  where  $\varphi < 1$  is the social marginal welfare weight for top-income earners (it measures the social value in euros of transferring a marginal euro to an income earner in the top-tax bracket) and  $\epsilon^c$  is the compensated elasticity of taxable income,  $\eta$  is the income elasticity, and  $a$  is the Pareto parameter of the earnings distribution. Equation is therefore an explicit formula for the optimal asymptotic top income tax rate if the social welfare weight is taken as exogenous. As in Diamond (1998), Dahan and Strawczynski (2012) rely on the direct limit argument i.e. the limit of optimal non-linear marginal tax rates when the wage tends to infinity<sup>9</sup>. They show that Diamond's asymptotic tax formula is not limited to the linear case and can be used for the non-linear utility of consumption as well. Dahan and Strawczynski (2012) also replicates the optimal asymptotic tax of Mirrlees (1971: equation 66) but in terms of labor supply elasticities instead of marginal utilities (see footnote 5 above).

## 2.3 The asymptotic solution

As noted by Mirrlees (1971) the asymptotic marginal tax rates are sensitive to the nonlinear utility of consumption. Above in (6) we assume that utility of consumption is linear. If the marginal utility of consumption takes the

---

<sup>9</sup>Most relevant cases of asymptotic tax rates were already analyzed by Mirrlees (1971, p. 188-193).

form of  $U_x = x^{-\mu}$  then the asymptotic marginal tax rates are sensitive to the value of  $\mu$  in the neighbourhood of 1. For example with a constant labour supply elasticity and  $U_x = x^{-2}$  the formula (9) in Saez (2001) yields the optimal asymptotic rate equal to 100 per cent (see also Dahan and Strawczynski (2012)). Mirrlees (1971) devoted much attention to the case where the utility function is of the Cobb-Douglas form,  $u = \log x + \log(1 - y)$ , the social welfare function is isoelastic, and skills are lognormally distributed or according to the Pareto distribution. The Marginal tax rate with the Pareto distribution tends asymptotically to  $\frac{2}{1+\theta}$  as  $n$  tends to infinity, where  $\theta$  is the shape parameter or Pareto parameter. This is also true for the Champernowne distribution (see Section 4). We give more complete description of the solution than in Mirrlees (1971, case ii p. 196-200) in the case of the Pareto or Champernowne distribution. When the path to the singular solution starts from  $y = 0$ , this implies that the marginal tax rates increase monotonically from  $t = \frac{1}{(1+\theta)}$  to  $t = \frac{2}{(1+\theta)}$ , where  $\theta$  is the shape or Pareto parameter in the Champernowne distribution (see the derivation in the appendix B). It is also important to note that these asymptotic results are independent of the net revenue requirement. Interestingly, the same asymptotic value holds - for the current skill distribution and preferences<sup>10</sup> - both in the Rawlsian or maximin and utilitarian cases as long as the marginal utility of consumption goes to zero as wage goes to infinity. In other words, although the shapes of the respective marginal tax schedules differ radically in the maximin and utilitarian cases at the lower end of the income distribution, they converge at the upper end.<sup>11</sup> This is not surprising. In the utilitarian case the weight attached to the top incomes tend to zero when  $n$  goes infinity. This is also in the case where the government minimizes some well-behaved poverty index. Of course, this convergence may not be apparent over the income range of practical interest. Whether this is the case, it will emerge in the numerical simulations provided below.

### 3 The taxable income elasticity at the top in Finland

Empirical work on the incentive effects of labour income taxation has generally identified quite low labour supply elasticities. Much of this research has analyzed labour supply in terms of hours. Compensated elasticities have often been found to be near 0.1 or 0.2, which implies that the elasticity of

<sup>10</sup>Mirrlees (1971) notes this in a footnote on the page 200.

<sup>11</sup>In fact this is also true in numerical simulations. Tuomala (2006) shows that both in maximin and utilitarian cases the marginal tax rates are almost the same at the 99 per cent point of the skill distribution.

substitution between leisure and consumption is around 0.5 in the CES utility function. It has been long recognised that behavioural responses of taxation are not confined to participation and hours worked. Feldstein (1995) proposed that we should examine the response of taxable income to changes in tax rates. Taxable income is determined not only by hours and participation. Individuals can respond to taxation on other margins such as job choice, intensity of work, timing of compensation, tax avoidance and tax evasion. Feldstein (1995) found very high elasticities exceeding one. Subsequent research generated considerably smaller estimates. In the recent survey on taxable income elasticities, Saez et al. (2012) conclude:

*“The most reliable longer-run elasticity estimates range from 0.1 to 0.4, suggesting that the U.S. top marginal rate is far from the top of the Laffer curve, but greater than one would calculate if the sole behavioural response was labour supply”.*

Much attention has been devoted to the effects of top marginal tax rates on the earnings distribution. As pointed out by Atkinson et al. (2009) higher top marginal tax rates can reduce top reported earnings through different channels. In particular, it has long been shown that the bulk of the elasticity response for top incomes comes from income shifting between various tax bases. For instance, lower capital income tax rates might lead to a rise in top taxable incomes reported as capital income, but this rise can be almost entirely offset by a corresponding decline in taxable earned income reported to the labour income tax. Jäntti et al. (2009) argue that the 1993 reform, which introduced dual income taxation, had an impact on the level and composition of top incomes<sup>12</sup>. Figure 1 documents a surge in capital incomes (mainly dividends) of top 1 per cent starts just after the 1993 reform. Figure 1 also shows that top tax rates on upper income earners have declined significantly after the 1993 reform. There is not much change in the composition of incomes among the next 4 per cent. The gap between the tax rates on labour and capital income was huge after the 1993 reform<sup>13</sup>. The dual income tax dropped the marginal tax rates on capital income the more the higher was the person’s total income before the reform. Those entrepreneurs who saw the largest reduction in the marginal tax rates on capital income also experienced the largest increase in capital income. At the same time, the share of labour income decreased for these taxpayers. The increase in

<sup>12</sup>The proportion of capital income was in 2007 53 per cent of income in the top one per cent group (Figure 1). It was 11 per cent in 1990.

<sup>13</sup>*“The (Finnish) system seems to offer generous opportunities for tax-avoidance by transforming labor income into capital income. For example, retained corporate profits will increase the amount that is taxed as capital income, and capital gains on shares are only subject to capital income tax”.* Lindhe et al. (2002).

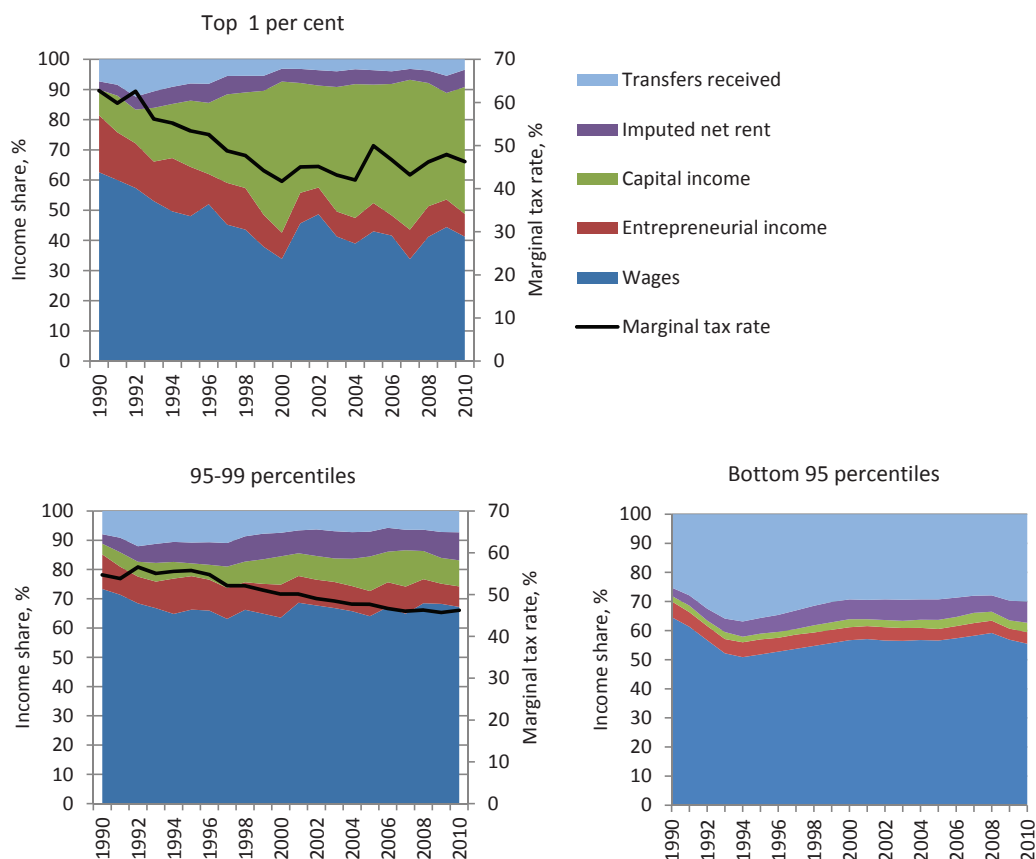


Figure 1: Marginal tax rate and gross income items for top 1 per cent, 95-99 percentile and bottom 95 percentiles

the share of capital income out of total income was much more modest for high-income employees.

The Finnish reform of 1993 sharply reduced the marginal tax rates on capital income to some corporate owners but did not change simultaneously the taxation of labour income. Corporate owners who can afford to save can reduce their intertemporal tax bill by taking less labour income out of the firm, which increases the net worth of the company. The increased value of net assets, in turn, increases the share of dividends that can be paid tax free in the future. People who are not owners of closely held corporations do not have an access to this route. Pirttilä and Selin (2010) examine whether the responses to the Finnish dual income tax reform of 1993 were different

among entrepreneurs and employees. The idea is that entrepreneurs have more leeway for income shifting than employees. They first make the tax base as constant as possible so that legislation changes governing the tax base would not distort our inference. They then estimate, using the approach in the elasticity of taxable income literature (for this, see e.g. Gruber and Saez (2002) and Aarbu and Thoresen 2001), how taxable capital and labour income reacted to changes in the marginal tax rates on labour and capital income.

In any case, the key point is that most of the behavioural response of top incomes to top tax rates seems to be due not to a real change in economic activity and output, but simply to a re-labelling of income outlays over various tax bases. Using the terminology introduced by Saez et al. (2012) in their survey, the behavioural response of top incomes involves substantial tax externalities which like all externalities have an impact on welfare and policy analysis. In general the literature estimates this elasticity based on the sum of labour and capital income. Top income shares together with information on marginal tax rates by income groups can be used to test theory and estimate the taxable elasticity.

The most dramatic changes in marginal income tax rates in Finland have taken place at the top percentile of the income distribution (Figure 2). Figure 11 in appendix C displays the average marginal tax rates and income shares of the next 4 per cent (income earners between the 95<sup>th</sup> and the 99<sup>th</sup> percentile). These figures show that the marginal tax rates by the top 1 per cent have declined since 1988<sup>14</sup>. It is interesting to note that the share received by the top 1 per cent of income recipients started to increase after 1993 not after 1989. In contrast to that for the next 4 per cent, the marginal tax rates of labour income and total income after 1993 differ only slightly from each other. This implies that income shifting possibilities are quite modest for this group.

Using the series displayed in Figures 1 and 2 and in appendix Figure 10 we can estimate the elasticity of taxable income around a tax reform episode taking place between pre-reform year and post-reform year as follows

$$\epsilon = \frac{(\log S_1 - \log S_0)}{\log(1 - t_1) - \log(1 - t_0)}, \quad (10)$$

where  $S_1$  the top 1 per cent income share after reform and  $S_0$  before reform,  $t_0$  the marginal tax rate (earned and capital) of the top 1 per cent before reform and  $t_1$  the marginal rate after reform.  $(1 - t)$  (or 1-MTR as in appendix D) is the net-of-tax rate. Here we assume that absent tax changes  $S_0 = S_1$ .

---

<sup>14</sup>See more closely the definition of the tax rates in Appendix C.

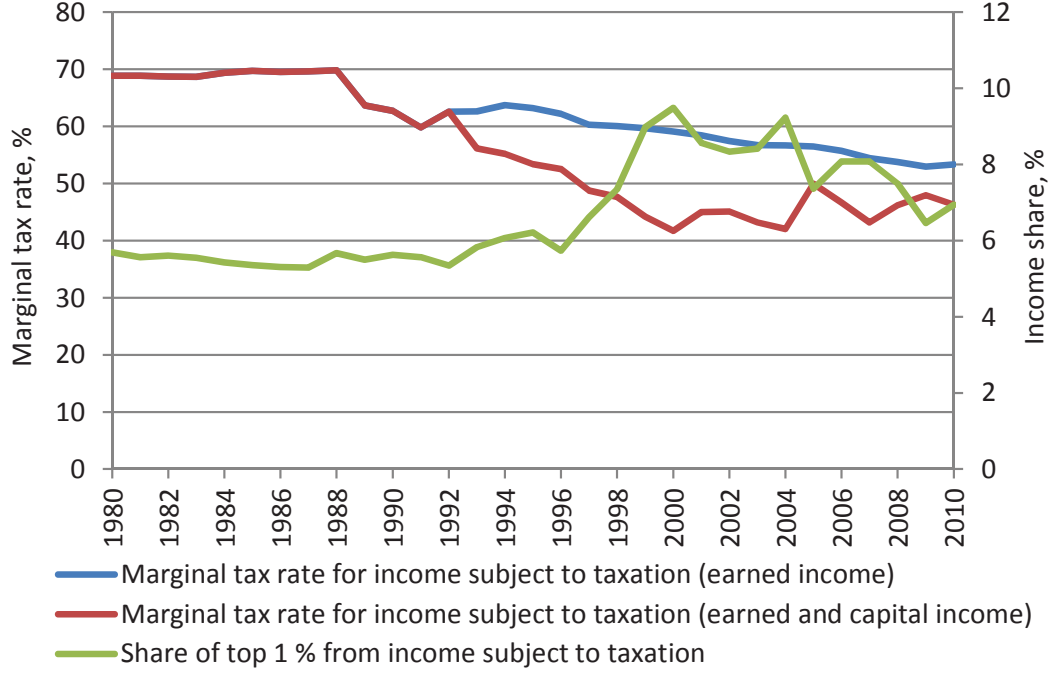


Figure 2: Marginal tax rates and income shares for top 1 per cent

Applying this simple method using series depicted in Figures 1, 2 and 10 above around the 1989/90 tax reform by comparing 1989 and 1991 generates an elasticity of 0.13 for the top 1 per cent (column 2).<sup>15</sup> Column 3 in Table 1 shows that the elasticities for the next 4 per cent. Comparing 1992 and 1995 around the tax reform 1993 gives a much larger elasticity of 0.69 for the top 1 per cent. It shows that the elasticity estimates obtained in this way are sensitive to a specific reform.

We also estimate the elasticity using the full time series evidence. We estimate the elasticity  $e$  with a log-form regression specification of the form:

$$\log(\text{Top Income Share}) = \text{constant} + e \log(1 - t) + \varepsilon \quad (11)$$

<sup>15</sup>The elasticity estimates in Pirttilä and Selin (2011) for the mean income fall in range 0.2-0.4.

**Table 1 Elasticity estimates using top (1 per cent) income share series in Finland 1980-2010<sup>a</sup>**

	Top 1 % DD <sup>b</sup> using the next 4 % as control (1)	Top 1 % Simple difference (2)	Top 95-99 % Simple difference (3)
A			
The tax reform episodes			
1989 vs. 1991 (Tax reform 1989/90)	0.32	0.13	0.01
1992 vs. 1995 (Tax reform 1993)	0.49	0.69	
B			
Full time series 1980-2010			
No time trends	1.31 (5.43)	0.67 (9.55)	0.06 (2.42)
Linear time trend	1.31 (5.30)	0.81 (4.59)	0.00 (0.03)

<sup>a</sup> t-values in the parentheses.

<sup>b</sup> DD = difference in differences.

Such a regression without time trend yields a very high estimate of the elasticity of 0.67 for the top 1 per cent. For the next 4 per cent the estimated elasticity is rather small 0.06. Inequality has also changed for the reason not related to taxation. To take into account these other considerations we could add some controls. This is not an easy task. We added time trends to (11). A combination of linear, square, and cube time trends did not improve the model. As pointed out by Saez et al. (2012) the problem with time-trend specification is that we do not know what time trend specifications are necessary for non-tax related changes.<sup>16</sup>

We also have what is referred to as a difference in differences estimates. These are presented in column (1) of table 1. For that we assume, absent the tax change, the top 1 per cent share would have increased as much as the next 4 per cent.<sup>17</sup> These estimates are rather high for the full time-series regression and they don't change with different time trends.

However there is a great deal of uncertainty around these numbers. The elasticity is estimated using changes to top incomes that happened during

<sup>16</sup>See also Piketty, T., E. Saez and S. Stantcheva (2011).

<sup>17</sup>

$$\epsilon = \frac{\log(S_1/S_1^k) - \log(S_0/S_0^k)}{\log[(1-t_1)/(1-t_1^k)] - \log[(1-t_0)/(1-t_0^k)]}$$

where  $t_k$  and  $S_k$  are the marginal tax rate and the income share for the control group, top



the 1990s, a period when the top rate of income tax was falling and when income inequality was increasing. This approach may then confuse responses to the policy with any underlying factor (e.g. income shifting<sup>18</sup>) increasing inequality, this could mean that our elasticity estimates are too high. This is especially true for the DD full time series regression. Namely income shifting was possible in practice for the top 1 per cent taxpayers. We should also bear in mind that the taxable income elasticity is not derived from immutable preferences, but are affected by the structure of tax system (see Kopczuk and Slemrod, 2002).

#### 4 Distribution of top incomes: Estimation of the Pareto and Champernowne-Fisk distributions

The excellent Pareto fit of the top tail of the distribution has been well known for over a century since the pioneering work of Pareto (1896) and verified in many countries and many periods, as summarized in Atkinson et al. (2011). The top tail of the income distribution is closely approximated by a Pareto distribution characterized by a power law density of the form  $f(z) = (1/z^{1+a})$  where  $a > 1$  is the Pareto parameter. Such distributions have the key property that the ratio  $z_m/z^*$  is the same for all  $z^*$  in the top tail and equal to  $b = \frac{a}{a-1}$ .  $z^*$  is the top  $x$  per cent threshold income and  $z_m$  is the average income of top  $x$  per cent. The higher  $a$  (i.e. lower coefficient  $\frac{a}{a-1}$ , (i.e. less fat upper tail) imply lower inequality, and conversely ( $b =$  inverted Pareto parameter). A lower coefficient means larger top income shares and higher income inequality. In Finland 1990-2010 the Pareto parameter (taxable income) varies between 3.7 (1994) and 1.79 (2004) (see Figure 3).

Figure 4 depicts the ratio  $z_m/(z_m - z^*)$ , with  $z^*$  ranging from zero to 300 000 euros annual incomes in 2000, 2005 and 2010. The ratios in different years show that top tail of the distribution is very well approximated by a Pareto distribution (see also Figure 4). Especially this is so in 2010. Then the ratio was 2.4. The top 1 per cent threshold  $z^*$  was 99 398 euros and  $z_m$  was 175 680 euros in 2010. In 1990 the top 1 per cent threshold started at 70 605 euros, and the average income of people in that bracket was 97 146 euros, then  $a = 3.66$ .

---

95-99 percentiles. And a log-form regression specification of the form:

$$[\log(\frac{S_1}{S_1^k}) - \log(\frac{S_0}{S_0^k})] = constant + e [\log(\frac{1-t_1}{1-t_1^k}) - \log(\frac{1-t_0}{1-t_0^k})] + \varepsilon$$

(see Appendix D).

<sup>18</sup>Figure 1 provides indirect evidence on income shifting within the top 1 per cent.

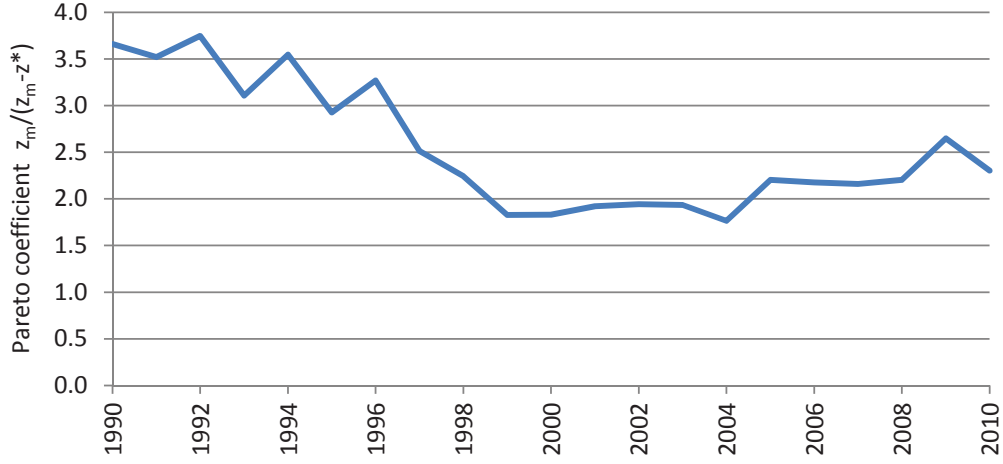


Figure 3: Empirical Pareto coefficient,  $a$ , in Finland 1990-2010.  $z^*$  is the top 1 per cent threshold income and  $z_m$  is the average income of top 1 per cent

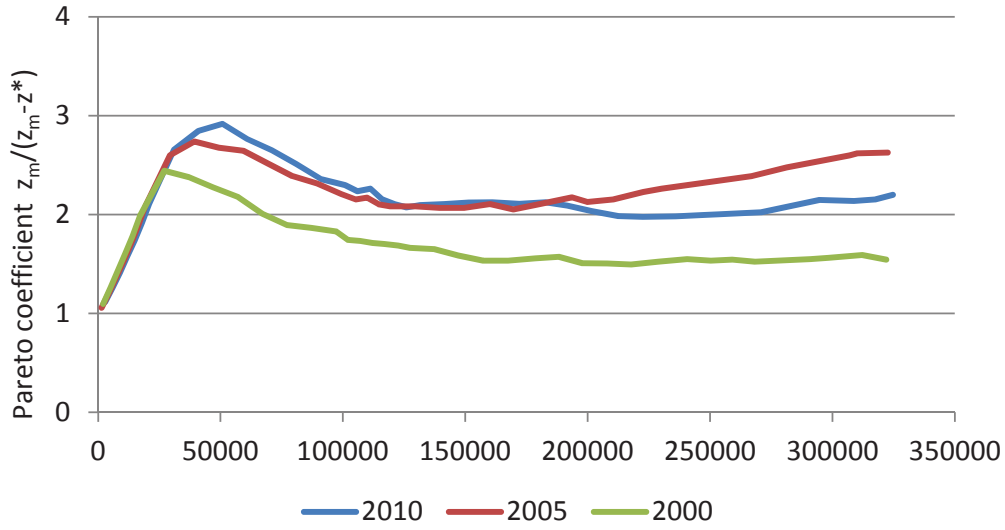


Figure 4: Pareto coefficient  $a = z_m/(z_m - z^*)$

We now consider an alternative to the Pareto distribution. Champernowne (1952) proposes a model in which individual incomes are assumed to follow a random walk in the logarithmic scale. Here we use the two parameter version of the Champernowne distribution. This distribution approaches asymptotically a form of Pareto distribution for large values of wages but it also has an interior maximum. As for the lognormal, the Champernowne

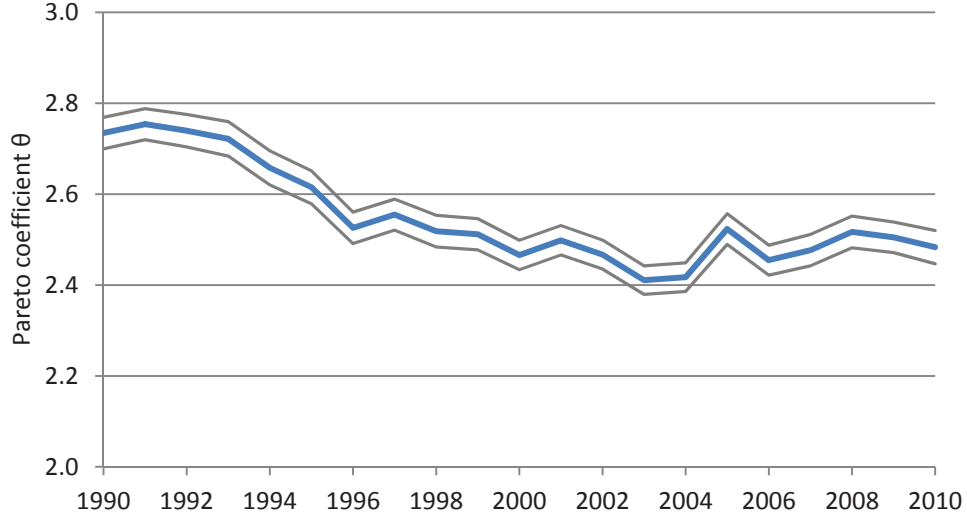


Figure 5: The shape parameter  $\theta$  (with confidence interval): Champernowne-Fisk distribution

distribution exhibits the following features: asymmetry, a left humpback and long right-hand tail; but it has a thicker upper tail than in the lognormal case. The probability density function of the Champernowne distribution is

$$h(z) = \theta \left[ \frac{m^\theta z^{\theta-1}}{(m^\theta + z^\theta)^2} \right] \quad (12)$$

in which  $\theta$  is a shape parameter and  $m$  is a scale parameter. The cumulative distribution function is

$$H(z) = 1 - \frac{m^\theta}{m^\theta + z^\theta} \quad (13)$$

For the distribution or Mills ratio:

$$\lim_{n \rightarrow \infty} \frac{1 - H(z)}{z h(z)} = \lim_{n \rightarrow \infty} \frac{m^\theta + z^\theta}{\theta z^\theta} \rightarrow \frac{1}{\theta}. \quad (14)$$

Equation (14) confirms that the Champernowne distribution approaches asymptotically a form of Pareto distribution for large values of wages.

Based on Finnish income distribution data (cross section) we estimated by using maximum likelihood methods several two and three parameter distributions with corresponding measures of goodness of fit (several of them plus the log-likelihood value for the estimated model). Among two parameter distributions the Champernowne is the best fitting for pre-tax income distribution in Finland (1990-2010). The  $\theta$ -parameter varies from 2.78 to 2.4 (see Figure 5). Over the period from the latter part of 1990s to 2010

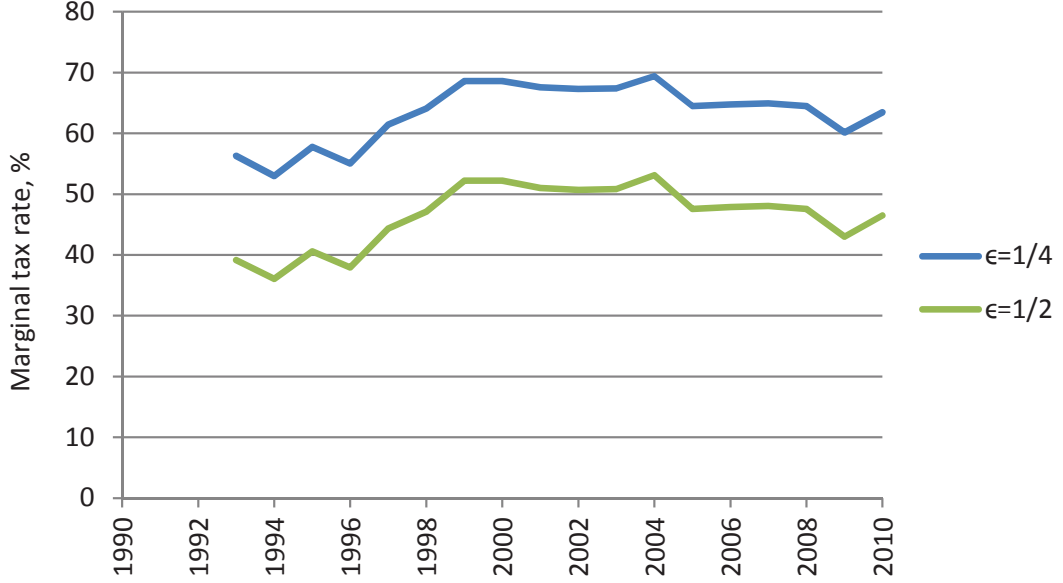


Figure 6: Optimal revenue maximizing top marginal tax rates (%) ( $\epsilon = \frac{1}{4}, \frac{1}{2}$ ) 1993-2010

the  $\theta$ -parameter was almost constant being around 2.5. Hence  $\theta = 2$  reflects a low range estimate (high inequality) and  $\theta = 3$  in turn a high range estimate (low inequality). The Gini coefficients estimated by this distribution ( $Gini = 1/\theta$ ) are quite close to those calculated from the data. Interestingly the location parameter  $m$  in our notations (median) in the Champnowne distribution is quite close to that calculated from the data.

## 5 Optimal top rates for the different assumptions in Finland

We employ the property of (8) to calculate optimal top income tax rates using reduced-form estimates of the taxable income elasticity for high incomes (in section 4) and a Pareto parameter of Figure 3 from the Finnish income distribution. Figure 6 displays our simple calculations based on the approximation formula (8) for two different values of  $\epsilon$  over the period 1993-2010. The first and second ones,  $\epsilon = 1/4$  and  $1/2$  are mid-range estimates in Table 1 and also in the recent survey by Saez et al. (2012).

Tables 2a and 2b illustrate with some parameter values top marginal tax rates with different social objectives (utilitarian and revenue maximizing).

The top tax rate was 44.3 percent in 2010. Maximizing revenue on the high earners implies that for example the current (2010) top rate in Finland has gone beyond the top of the Dupuit-Laffer curve if and only if  $\epsilon$  is larger than 0.55 given  $a = 2.3$ . In Figure 7 we show the corresponding values for elasticity from 1990 to 2010. Given our estimates on Pareto coefficients and elasticity, we are convinced that top tax rate varies between 50-67 %.

We can also consider the inverse problem of determining the social welfare weight parameter  $\varphi$  for a given top tax rate. Assuming the labour supply elasticity is equal to 0.5 (0.25) and Pareto parameter  $a$  as shown in Figure 3, then we can calculate the welfare weights  $\varphi$  implied by the top tax rates in Finland over the years 1990-2010 (Figure 8). For example in 2010  $\varphi$  is equal to 0.63 (0.39). The tax rates in Table 2a and b are based on the approximation formulas (8) and (9). They give smaller values than the asymptotic rates. For example in the revenue maximizing case with  $a = 2$  and  $\epsilon = 0.5$  we have  $t = 50$  % but the asymptotic rate with the same parameter values is 60 % (from formula (6)).

**Table 2a Utilitarian top marginal tax rates (%) for  $\varphi = \frac{1}{3}$  and  $\frac{1}{2}$**

	$a = 1.5$	$a = 2.0$	$a = 2.5$	$a = 3.0$
$\varphi = \frac{1}{3}$				
$\epsilon = 0.25$	64.1	57.3	51.7	47.2
$\epsilon = 0.50$	47.2	40.1	34.9	30.9
$\varphi = \frac{1}{2}$				
$\epsilon = 0.25$	57.1	50.0	44.4	40.0
$\epsilon = 0.50$	40.0	33.3	28.6	25.0

**Table 2b Rawlsian (maximizing revenue on the high earner) top marginal tax rates (%) for  $\varphi = 0$**

	$a = 1.5$	$a = 2.0$	$a = 2.5$	$a = 3.0$
$\varphi = 0$				
$\epsilon = 0.25$	72.7	66.7	61.5	57.1
$\epsilon = 0.50$	57.1	50.0	44.4	40.0

Responses to tax rates can also take the form of tax avoidance. Tax avoidance can be defined as changes in reported income due to changes in the form of compensation but not in the total level of compensation. In the Finnish tax system tax avoidance opportunities arise when taxpayers can shift part of their taxable labour income into capital income. As mentioned above income shifting was possible in practice for the top 1 per cent of taxpayers i.e. for

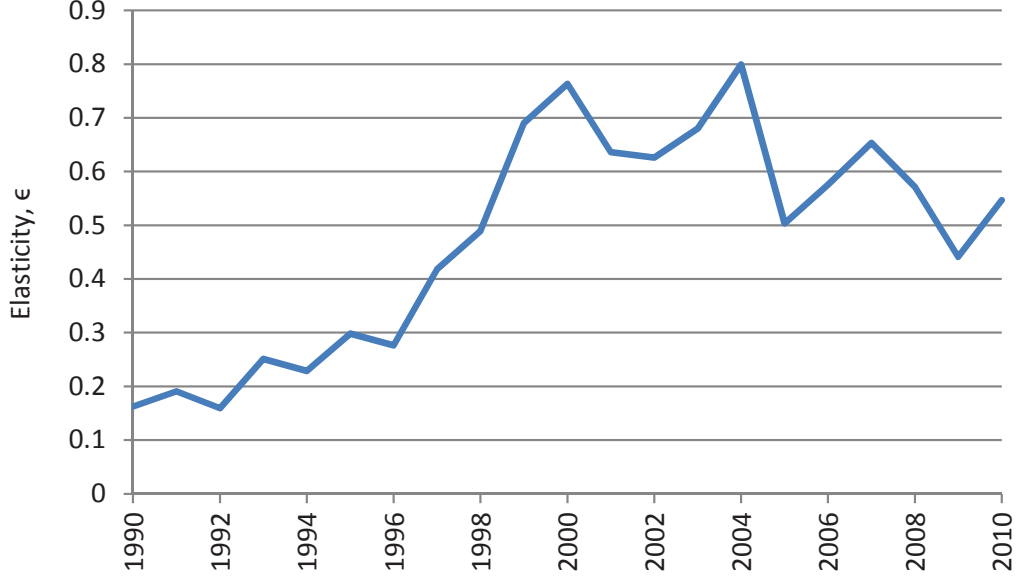


Figure 7: Revenue maximizing elasticity ( $\epsilon$ ) in Finland 1990-2010

wealthy owners of closely held companies. Piketty et al. (2012) extend the revenue maximizing top tax rate formula with the elasticity of tax avoidance. The formula takes the following form  $t = \frac{(1+\tau a \epsilon_A)}{1+a\epsilon}$  (see Piketty et al. 2012 for a derivation), where  $\epsilon_A = \frac{(1-t)dr}{zd(t-\tau)}$  is the tax avoidance elasticity ( $r$  is shifted income so that ordinary taxable income is a difference between real income and  $r$ ) and shifted income is taxed at a constant and uniform marginal tax rate  $\tau$  lower than  $t$ . For given  $\tau = 25\%$ <sup>19</sup>, we assume  $\epsilon = 0.5$  ( $0.7$ )<sup>20</sup>,  $\epsilon^* = 0.3$  ( $0.1$ ),  $\epsilon_A = 0.2$  ( $0.6$ )<sup>21</sup>,  $a = 2$  and  $\tau = 25\%$  then we obtain a revenue maximizing top rate  $t = 55\%$  ( $54.2\%$ ). What are tax policy implications of top tax rates based on this formula with the avoidance elasticity? They simply reflect weaknesses of the tax system. In fact the optimal tax system should minimize the income shifting channels.

For Pareto distributions,  $\frac{1-H(z)}{zh(z)}$  is constant. However, the empirical

<sup>19</sup>It was capital income tax rate in the 1993 reform).

<sup>20</sup> $e = (d/z)e^* + e_A$  where  $e^*$  is real labour supply elasticity, and the ordinary taxable income  $z = d - r$  ( $d$  = real income).

<sup>21</sup>We chose the values of elasticities assuming that in Table 1 the elasticity based on the 1993 reform reflects also income shifting but not the estimate of elasticity based on the 1989 reform.

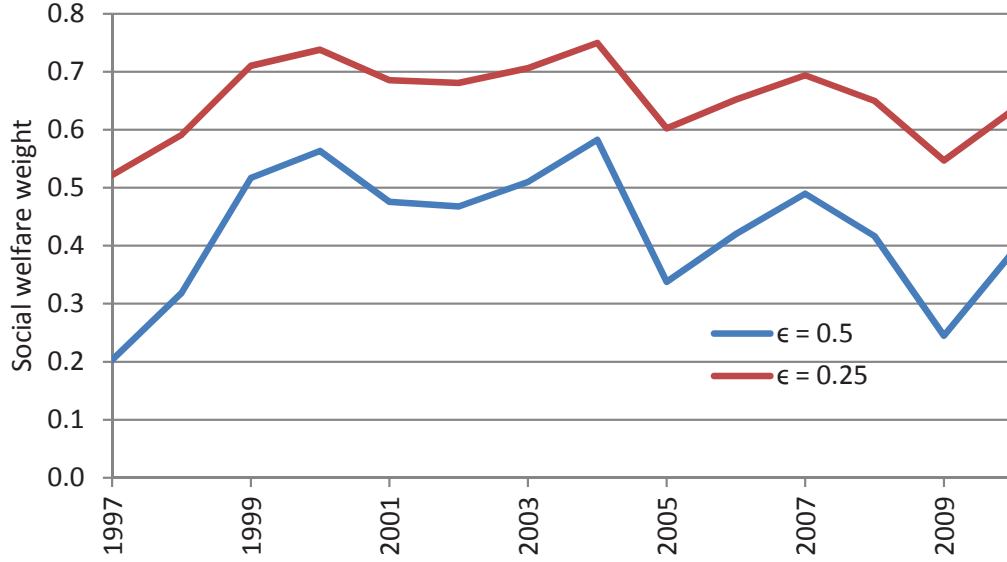


Figure 8: Social welfare weight parameter  $\varphi$ , 1997-2010

Finnish income distribution is not a Pareto distribution at lower income levels. Therefore we adopt the Champnowne distribution. Assume an elasticity of 0.25, the social marginal welfare weight  $\varphi = 0.5$  at the income level corresponding to the top decile threshold is 0.5 where  $\theta = 2.5$  i.e.  $\frac{1-H(z)}{zh(z)} = 0.51$  (2010) and  $\frac{1-F(n)}{nf(n)} = 0.45$ . The optimal utilitarian marginal tax rate at this income level is 51 per cent.

### The nonlinear utility of consumption and the asymptotic marginal tax rate

The special cases considered in the previous section yield insights but within the framework of the assumptions made. How robust are these insights? What happens when we move away from quasi-linearity? As we mentioned in section 2 Mirrlees (1971) found that the asymptotic marginal tax rates are sensitive to the nonlinearity of utility of consumption. We focus here on an example with the Cobb-Douglas preferences and the Champnowne

distribution. In table 3 we compare asymptotic rate with the optimum nonlinear tax rate (numerical) at the upper part of  $n$ -distribution.

**Table 3 Asymptotic and numerical top marginal tax rates**  
 $u = \ln x + \ln(1 - y), f(n) = \text{Champernowne}$

	$\theta = 2$	$\theta = 2.5$	$\theta = 3$	$F(n)$
Asymptotic rate	67	57	50	
Numerical rate <sup>20</sup>	50.1	40.4	33.8	99

<sup>20</sup>Tuomala (2006).

A number of interesting points emerge from the patterns in Table 3. There are quite big differences between the numerical and asymptotic rates. The asymptotic solution is not a very accurate approximation for even top 1 per cent with  $\theta = 2, 2.5$  and 3. But this difference is much smaller than in Mirrlees (1971). The reason is that here we are using the distribution with a much thicker upper tail than that one used in Mirrlees (1971). The asymptotic rates are in line with those top rates we find in many advanced countries. It is perhaps slightly surprising to find rather high tax rates with the Cobb-Douglas preferences where the labour elasticity is rather high around 0.5 at the income level of top one per cent. In the case  $\theta = 2.5$  (2.0) the asymptotic rate is 57 (67) per cent. The asymptotic tax of Mirrlees (1971: equation 66) with the Pareto parameter equal to 2.0 is 60 per cent but his formula is not in terms of labor supply elasticities instead of marginal utilities (see footnote 5 above).

## 6 Concluding remarks

The current (2010) top marginal rate for top 1 per cent in Finland is 44.3 per cent. The corresponding top tax rate was 62.7 per cent in 1990. With the Pareto parameter  $a = 2.5$ , (2.0) the labour supply elasticity  $\epsilon = 0.25$  yields revenue maximizing top marginal tax rate  $t$  of 61.5 per cent (66.7 per cent) and utilitarian with positive welfare weight (1/3) yields  $t$  of 51.7 (57.3) per cent. The more general case with the Cobb-Douglas preferences and  $a = 2.5$  or  $\theta = 2.5$  gives the top rate of 57 per cent. It is clear that this kind of calculations of the optimal tax rate needs to recognize explicitly that there is much uncertainty particularly related to supply elasticities and not only to rely on one estimate. Notwithstanding one conclusion we can draw from our application without any uncertainty is that the current top marginal tax rate is not close to the top of the Dupuit-Laffer curve.



How to assess our results for top marginal tax rate depends also on whether elements left out of the model change them. There are good reasons to suspect that the labour market of top income earners deviates from the standard competitive model in a number of important respects. Persson and Sandmo (2005) investigate optimal income taxation in a “tournament” model where wages are determined not by productivity but by one’s productivity relative to other workers. As they note, such a model might be particularly relevant to the salaries of top executives. Would it therefore be a more suitable framework within which to examine the optimal top tax rate? There are problems with the tournament explanation, however, and there is no real evidence that it applies to executive pay. For example, tournaments might provide poor incentives when it is apparent that one player is likely to win and others likely to lose the competition (due to differences in skills or other qualities) (see Eriksson, 1999).

As suggested by Simula and Trannoy (2010) taking potential losses of tax base due to migration into account can significantly reduce the level of the optimal top marginal tax rate. In a simple case this can be seen by extending formula (6) adding migration effects using migration elasticity  $\eta$ . It is the elasticity of numbers with respect to after-tax income. We have the following formula  $t = \frac{1}{1+a\epsilon+\eta}$ . From Table 2b we see that when  $a=2$  and  $\epsilon=0.25$ ,  $t=2/3$ . If there is migration with elasticity  $\eta = 0.4$ , then the revenue maximizing tax rate decreases to 52.6. Thus, large migration elasticities could indeed decrease significantly the ability of countries to tax high incomes. We should emphasize that this is a single-country optimum. A single country does not recognize the external effects it might impose on other countries by cutting its top tax rate. A global welfare perspective and with complete fiscal coordination, the migration elasticity is not that relevant for optimal tax policy.

Finally, if the economic activity at high incomes is primarily socially unproductive rent-seeking, then it would be plausible to impose higher marginal rates at top income levels than those calculated above.

## References

- Aaberge, R., and Colombino, U. (2006), Designing Optimal Taxes with a Microeconomic Model of Household Labour Supply, *Discussion Papers* 475), Research Department of Statistics Norway.
- Aarbu, K.O., and Thoresen, T.O. (2001), Income Responses to Tax Changes-

- Evidence from the Norwegian Tax Reform, *textitNational Tax Journal* 54, 319-338.
- Atkinson, A.B. (1990), Public Economics and the Economic Public, *European Economic Review* 34, 225-248.
- Atkinson, A.B. (2012), The Mirrlees Review and the State of Public Economics, *Journal of Economic Literature* 2012 3/50, 770-780.
- Atkinson, A.B., T. Piketty and Saez, E. (2009), Top Incomes and the Long-run History of Inequality, in A.B. Atkinson and T. Piketty (eds), *Top Incomes over the Twentieth Century* vol. II, Oxford University Press.
- Atkinson, A.B., and Brandolini, A. (2010), On Analyzing the World Distribution of Income, *World Bank Economic Review* 1/24, 1-37, Oxford University Press.
- Atkinson, A.B., T. Piketty and Saez, E. (2011), Top Incomes in the Long Run of History, *Journal of Economic Literature* 1/49, 3-71, American Economic Association.
- Bevan, D., (2005), On the Shape of the Optimal Tax-transfer Schedules under Non-welfarist Objectives, mimeo, Oxford University.
- Champernowne, D.G., (1952), The Graduation of Income Distributions, *Econometrica* 20, 591-615.
- Cowell, F., (1977), *Measuring Inequality*, Phillip Allen, London.
- Dahan M., and Strawczynski, M. (2000), Optimal Income Taxation: An Example with a U-shaped Pattern of Optimal Marginal Tax Rates: Comment. *American Economic Review* 3/90, 681-686.
- Dahan, M., and Strawczynski, M. (2012), Optimal Asymptotic Income Tax Rates, *Journal of Public Economic Theory, Association for Public Economic Theory* 5/14, 737-755.
- Diamond, P. (1998), Optimal Income Taxation: An Example with a Ushaped Pattern of Optimal Marginal Tax Rates, *American Economic Review* 1/88, 83-95.
- Eriksson, T. 1999, Executive Compensation and Tournament Theory: Empirical Tests on Danish Data, *Journal of Labor Economics* 2/17, University of Chicago Press, 262-280.

- Feldstein, M. (1995), The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act, *Journal of Political Economy* 103, 551-572.
- Feldstein, M. (2012), The Mirrlees Review, *Journal of Economic Literature* 2012 3/50, 781-790.
- Fisk, P.R. (1961), The Graduation of Income Distribution, *Econometrica* 29, 171-185.
- Goolsbee, A. (1998), It's not about the Money: Why Natural Experiments don't Work on the Rich, *NBER Working Paper* 6395.
- Gruber, J. and Saez, E. (2002), The Elasticity of Taxable Income: Evidence and Implications, *Journal of Public Economics* 84, 1-32.
- Jäntti, M., M. Riihelä, R. Sullström and Tuomala, M. (2009), Trends in Top Income Shares in Finland, in A.B. Atkinson and T. Piketty (eds.) *Top Incomes around the World vol. II*, Oxford University Press.
- Kanbur, R., and Tuomala, M. (1994), Inherent Inequality and the Optimal Graduation of Marginal Tax Rates, *Scandinavian Journal of Economics* 2/96, 275-282.
- Kanbur, R., and Tuomala, M. (2011), Charitable Conservatism, Poverty Radicalism and Good Old Fashioned Inequality Aversion, *Journal of Economic Inequality* 3/9, 417-431.
- Kopczuk, W., and Slemrod, J.B. (2002), The Optimal Elasticity of Taxable Income, *Journal of Public Economics* 84, 91-112.
- Landais, C. (2007), Les Hauts Revenus en France (1998-2006): Unexplosion des Ingalits? Paris School of Economics, mimeo.
- Lindhe, T., J. Södersten, and Öberg, A. (2002), Economic Effects of Taxing Closed Corporations under a Dual Income Tax, *Ifo Studies* 48, 575-610.
- Mirrlees, J. (1971), An Exploration in the Theory of Optimal Income Taxation, *Review of Economic Studies* 2/38, 175-208.
- Pareto, V., (1896, 1897), Cours d'économie politique profess a l'universit de Lausanne, Vol. I, 1896; Vol. II, 1897.
- Persson, M., and Sandmo, S. (2005), Taxation and Tournaments, *Journal of Public Economic Theory* 7, 543-559.

- Piketty, T., E. Saez, and Stantcheva, S. (2011), Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities, *NBER Working Papers 17616*, National Bureau of Economic Research, Inc.
- Pirttilä, J. and Selin, H. (2010), Income Shifting within a Dual Income Tax System: Evidence from the Finnish Tax Reform of 1993, *Scandinavian Journal of Economics* 1/113, 120-144.
- Rawls, J. (1971), *A Theory of Justice*, Harvard University Press, Cambridge Mass.
- Rawls, J. (1982), Social Unity and Primary Goods, in *Utilitarianism and Beyond*, eds. by A. Sen and B. Williams, 159-185, Cambridge Press.
- Riihelä, M., R. Sullström, and Tuomala, M. (2005), Trends in Top Income Shares in Finland, Government Institute for Economic Research, *Discussion Papers 371*, Helsinki.
- Riihelä, M. (2009), Essays on Income Inequality, Poverty and the Evolution of Top Income Shares, Government Institute for Economic Research, *Publications 52*, Helsinki.
- Röed, K. and Ström, S. (2002), Progressive Taxes and the Labour Market: Is the Trade-off between Equality and Efficiency Inevitable? *Journal of Economics Surveys* 1/16.
- Sadka, E. (1976), On Income Distribution, Incentive Effects and Optimal Income Taxation, *Review of Economic Studies* 2/43, 261-268.
- Seade, J. K. (1977), On the Shape of Optimal Tax Schedules, *Journal of Public Economics* 1/7, 203-236.
- Saez, E. (2001), Using Elasticities to Derive Optimal Income Tax Rates, *Review of Economic Studies* 68, 205-229.
- Saez, E., J. Slemrod, and Giertz, S. H. (2012), The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review, *Journal of Economic Literature, American Economic Association* 1/50, 3-50.
- Simula, L. and Trannoy A. (2010). Optimal Income Tax under the Threat of Migration by Top-income Earners,” *Journal of Public Economics, Elsevier, vol. 94(1-2)*, 163-173, February.
- Slemrod, J. (1998), Methodological Issues in Measuring and Interpreting Taxable Income Elasticities, *National Tax Journal* 51, 773-788.

- Tuomala, M. (1984), On the Optimal Income Taxation: Some further Numerical Results, *Journal of Public Economics* 23, 351-366.
- Tuomala, M. (1985), Simplified Formulae for Optimal Linear Income Taxation, *Scandinavian Journal of Economics* 4/87, 668-672.
- Tuomala, M. (1990), Optimal Income Tax and Redistribution, Oxford University Press, New York and Oxford.
- Tuomala, M. (2006), On the Shape of Optimal Non-Linear Income Tax Schedule, Tampere Economic Working Papers Net Series, *Working paper* 49, (<http://tampub.uta.fi/econet/wp49-2006.pdf>).
- Tuomala, M. (2010), On Optimal Non-linear Income Taxation: Numerical Results Revisited, *International Tax and Public Finance* 3/17, 259-270.

## Appendix A Derivation of the equation (4)

The quasi-linear utility function  $u = x + V(1 - y)$  is defined over consumption  $x$  and hours worked  $y$ , with  $U_x = 1$  and  $V_y < 0$  (subscripts indicating partial derivatives). Introducing multipliers  $\lambda$  and  $\mu(n)$  for the budget constraint (3) and incentive compatibility constraint  $\frac{du}{dn} = -\frac{yV_y}{n} = g$  and integrating by parts, the Lagrangean becomes

$$L = \int_0^\infty ((G(u) + \lambda(ny - x))f(n) - \mu'u - \mu g)dn + \mu(\infty) - \mu(0)u(0). \quad (A1)$$

With quasi-linear preferences differentiating of the Lagrangean (A1) with respect to  $u$  and  $y$  gives the first-order conditions

$$L_u = (G' - \lambda)f(n) - \mu'(n) = 0 \quad (A2)$$

$$L_y = \lambda(n + V_y)f(n) + \mu(n)\frac{(V_y + yV_{yy})}{n} = 0 \quad (A3)$$

(A2) satisfies the transversality conditions

$$\frac{\partial L}{\partial u(0)} = \mu(0) = 0; \quad \frac{\partial L}{\partial u(\infty)} = \mu(\infty) = 0.$$

Integrating in (A2)

$$\mu(n) = \int_n^\infty [\lambda - G']f(m)dm. \quad (A4)$$

This latter satisfies the transversality conditions  $\mu(0) = \mu(\infty) = 0$  [integrating in (A2)]

$$\int_n^\infty \frac{d\mu}{dn}dn = \mu(\infty) - \mu(n).$$

The transversality conditions and (A3) imply  $\mu(n) > 0$ , for  $n \in (0, \infty)$ .

The transversality condition  $\mu(n) = 0$  yields  $\lambda = \int_0^\infty G'[U(x)]f(p)dp$ . From the first order conditions of government's maximization, we obtain the following condition for optimal relative marginal tax rate  $t(z)$ ; [Note:  $\frac{t}{1-t} = \frac{1}{1-t} - 1 = \frac{U_x n}{V_y} - 1$ ]

$$\frac{t}{1-t} = (1 + \frac{yV_{yy}}{V_y}) \frac{1}{\lambda n f(n)} \int_n^\infty [\lambda - G']f(m)dm. \quad (A5)$$

Multiplying and dividing (A5) by  $(1 - F(n))$  and with  $u = x - y^{1+\frac{1}{\epsilon}}/(1 + \frac{1}{\epsilon})$  we obtain the formula (6) in the text.

We can also translate the analysis from  $n$ -space to  $z$ -space. In the linearized income tax system when  $z(n) = ny(n(1-t))$ ,  $dz/dn = y + (1-t)ndy/d(n(1-t)) = n(1+\epsilon)$ . Let  $H(z)$  be the distribution function of households by income  $z$  (which equals  $ny$  and so is endogenous), with density  $h(z)$ . We have  $f(n) = H(z)y(1+\epsilon)$ . Hence and we can write (5A)

$$\frac{t}{1-t} = \frac{1}{\epsilon} \frac{(1-H(z))}{zh(z)} (1-G') \quad (\text{A6})$$

where

$$G' = \int_z^\infty G' h dz' / (1-H(z)). \quad (\text{A7})$$

## Appendix B An asymptotic solution of optimal tax problem

Now we consider a more general case with  $u = U(x) + V(1-y)$ . Define

$$v = \frac{1 + V_y/nU_x}{V_y + yV_{yy}}, \quad (\text{B1})$$

and rewrite (A3)

$$n^2 f v = \int_n^\infty \left( \frac{1}{U_x} - \frac{G'}{\lambda} \right) f(n) dn = 0. \quad (\text{B2})$$

Differentiating (B2) with respect to  $n$  and we have

$$\frac{dv}{dn} = -\frac{v}{n} \left( 2 + \frac{nf'}{f} \right) + \frac{1}{n^2} \left( \frac{G'}{\lambda} - \frac{1}{U_x} \right) \quad (\text{B3})$$

(B3) and the incentive compatibility constraint

$$\frac{du}{dn} = -\frac{yV_y}{n} \quad (\text{B4})$$

two differential equations in  $u$  and  $v$ , provide the solution to the optimal income tax problem, together with the condition (B2) and at  $n = n_0$ <sup>22</sup>

$$n_0^2 f(n_0) v_0 = \int_0^{n_0} \left( \frac{1}{U_x} - \frac{G'}{\lambda} \right) f(n) dn \quad (\text{B5})$$

---

<sup>22</sup> $n_0$ , largest  $n$  for which  $y(n) = 0$ , may be in some cases rather large. In the interval  $[0, n_0]$ ,  $y = 0$  and  $x = x_0$  and then  $u = U(x_0) - V(0)$ .

and the transversality condition  $\mu(\infty) = 0$  and (B2) require that

$$\lambda n^2 f v \rightarrow 0 \quad (n \rightarrow \infty). \quad (\text{B6})$$

The condition (B5) guarantees an accurate value for  $n_0$ .

We analyze asymptotic marginal tax rates in the following case: The utility function is  $u = \log x + \log(1 - y)$ . We denote  $s = 1 - y$  and put  $w = \log n$ . Now we can write

$$\lambda v = s(s - \frac{x}{n}). \quad (\text{B7})$$

For any social welfare function  $G$  with a property that  $\lim_{u \rightarrow \infty} G'(u) = 0$  we can simplify (B3)<sup>23</sup> and if  $f(n)$  is the Champernowne distribution, then (B3) becomes

$$\frac{dv}{dt} = -v[1 - \frac{\theta(1 - (\frac{m}{n})^\theta)}{(1 + (\frac{m}{n})^\theta)}] - s(1 - \frac{v}{s^2}). \quad (\text{B8})$$

Since

$$\lim_{n \rightarrow \infty} \frac{nf'}{f} = -(1 + \theta) \quad (\text{B9})$$

we can rewrite

$$\frac{dv}{dw} = v(\theta - 1) - s(1 - \frac{v}{s^2}) \quad (\text{B10})$$

Hence from (B4) and noting that  $u = \log n + \log(s - \frac{v}{s}) + \log s$  we have

$$\frac{du}{dw} = \frac{du}{dn} \frac{dn}{dw} = 1 - \frac{1}{s} \frac{ds}{dw} + \frac{[(1 + v/s^2) \frac{ds}{dw} - \frac{1}{s} \frac{dv}{dw}]}{(s - v/s)} = \frac{1 - s}{s}. \quad (\text{B11})$$

Using (B11) we can write

$$\frac{ds}{dw} = \frac{v(1 + \theta) - 2s^2}{2s}. \quad (\text{B12})$$

Denote  $\frac{v}{s^2} = t$  i.e. the marginal tax rate. It follows that

$$\frac{ds}{dw} = \frac{s}{2}[t(1 + \theta) - 2] \quad (\text{B13})$$

Differentiating  $t = \frac{v}{s^2}$  with respect  $w$  and substituting  $\frac{dv}{dw}$  from (B10) we obtain

$$\frac{dt}{dw} = t\theta - t - \frac{1}{s}(1 - t) - 2t \frac{ds}{dw} \frac{1}{s}. \quad (\text{B14})$$

---

<sup>23</sup>Equivalently we have a maximin case.



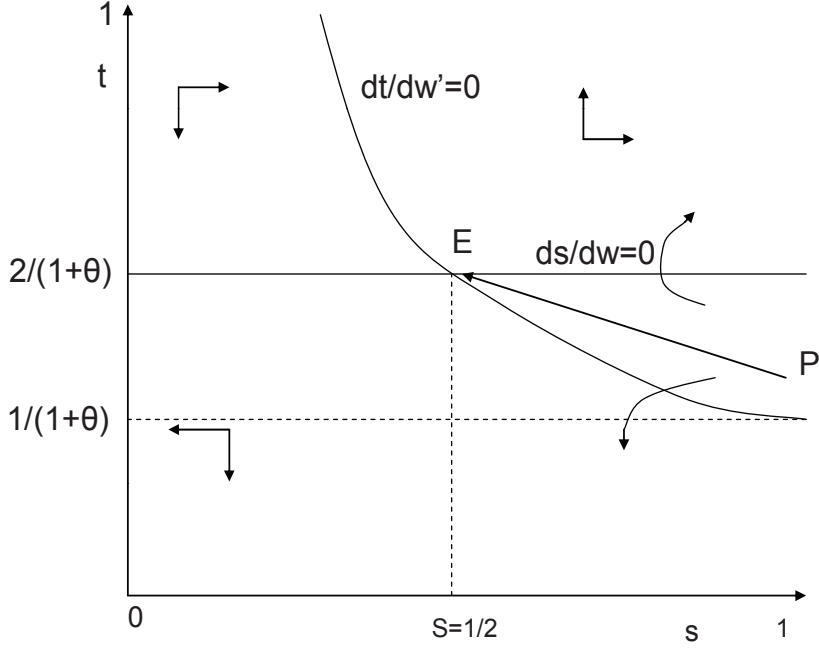


Figure 9: The solution to equation B13 and B15

Substituting to (B14) (B12) we have

$$\frac{dt}{dw} = (1 - t)[t(1 + \theta) - \frac{1}{s}]. \quad (\text{B15})$$

The solution to equation (B13) and (B15) is shown in Figure 8. The equations are autonomous in the sense that the evolution of  $t$  and  $s$  or  $(1 - y)$  is independent of  $n_0$ . Hence the solution of  $t$  is determined solely from (B13) and (B15). For  $\frac{dt}{dw} = 0$ , it has to be  $t = \frac{1}{(1+\theta)s}$  and for  $\frac{ds}{dw} = 0$  in turn  $t = \frac{2}{1+\theta}$ . Hence we have a complete description of the solution in the case of the Champernowne distribution. When the path to the singular solution starts from  $s = 1$ , this implies that the marginal tax rates increase monotonically from  $t = \frac{1}{1+\theta}$  to  $t = \frac{2}{1+\theta}$ .

## Appendix C

### Data and income concept

The top income share estimates from 1980 to 1995 are based on the Tabulated Income Data from Tax Tables (see Jäntti et al. 2010) and from 1996 to 2010 estimates based on the Income Distribution Statistics (IDS). The IDS data is a representative national sample. The concept of income is income subject to taxation (referring in the text to taxable income). Income subject to taxation also includes the deductions which are excluded from taxable income. Effective marginal earned income tax rates are estimated with a micro simulation model (JUTTA). Capital income tax rates are estimated from the IDS data. We have restricted the data to cover member of the population over 14 years of age in order to get two comparable databases.

Since 1989 several tax reforms have been implemented concerning personal taxation in Finland. The tax rates were reduced and the tax base was broadened. Furthermore, in 1990 the full imputation system of corporation taxes (*avoir fiscal*) was adopted to remove the double taxation of dividends. Since 1993 Finland has applied the dual income tax system where earned income is taxed at a progressive tax rate and capital income at a flat tax rate. The corporation tax rate and capital income tax rate were the same. The tax rate was 25 per cent from 1993 to 1995, 28 per cent from 1996 to 1999 and 29 per cent from 2000 to 2004. In our analysis incidence of capital taxation is straightforward: the shareholder also pays the corporation tax of dividends. In the case of the *avoir fiscal* both capital income and earned income were single-taxed.

The *avoir fiscal* tax system was abolished in 2005 and Finland switched to the partial double taxation of dividends. At the same time the capital tax rate was reduced to 28 per cent and the corporation rate to 26 per cent. To avoid the double taxation of dividends, part of capital income became tax free under personal taxation: 70 per cent of dividends were included in the shareholders' capital income. Furthermore, for owners of privately held businesses, dividends up to 90 000 euros (and dividends that were under 9 per cent of net wealth) were made tax exempt under individual taxation. As a result of the reform the average effective capital income tax rates increased for the top 1 per cent incomes in 2005.

The average effective marginal tax rates for the top 1 per cent incomes are illustrated in Figure 10. Since 1996, the effective marginal tax rate for earned income has decreased 10 per cent. At the same period average marginal tax rate for capital income has increased. Variation after 2004 comes from the fluctuation of the tax free part in capital income.

Since 1993 the total marginal tax rate is the weighted average of earned income and capital income. The total average tax rate for the top 1 per cent

decreased until 2000 because of the substantial difference between the earned income and capital income marginal tax rates and considerable increase in the capital income share (see Figure 1). The decrease of corresponding marginal tax rates for the next 4 per cent incomes has been more modest and it has stayed close to the marginal tax rate of earned income (see Figure 11).

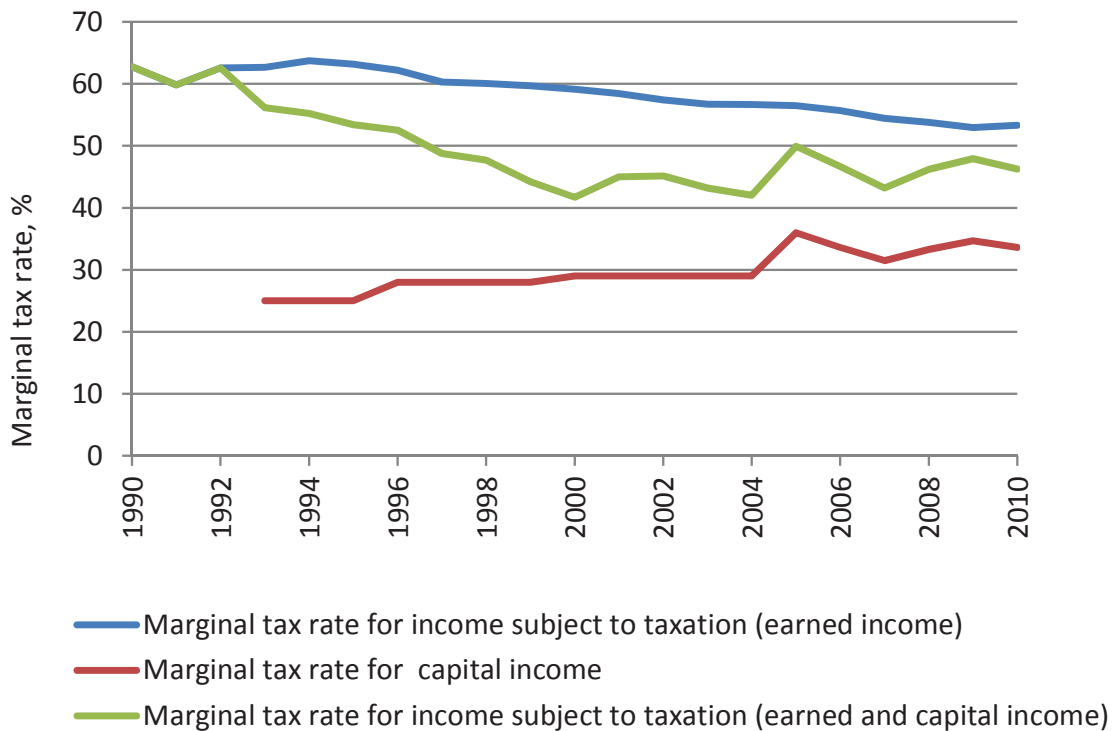


Figure 10: Marginal tax rates for top 1 percentiles

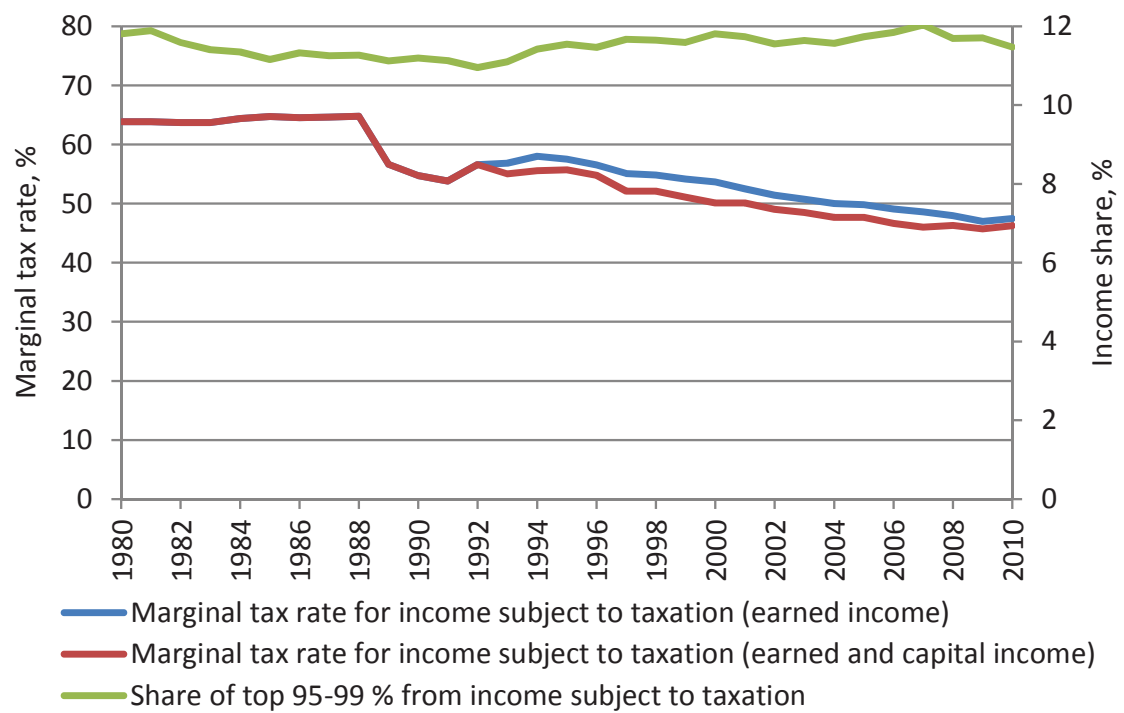


Figure 11: Marginal tax rates and income shares for top 95-99 percentiles

## Appendix D

**Table D1 Elasticity estimates**

Models	$a$	$e$	$b_1$	$b_2$	$b_3$	$R^2$	$D - W$	$S.E.E.$
Top 1 %								
No time trend	-0.938 (-34.9)	0.671 (9.5)				0.759	0.507	0.042
Linear time trend	-0.854 (-8.71)	0.817 (4.59)	-0.002 (-0.90)			0.765	0.510	0.042
Linear square trend	-0.718 (-6.06)	0.996 (5.10)	-0.011 (-2.10)	0.000 (1.88)		0.792	0.597	0.040
Cube time trend	-1.033 (-7.14)	0.289 (1.02)	-0.033 (-3.91)	0.003 (3.34)	0.000 (-3.10)	0.849	0.930	0.035

$$\log(S_{1,t}) = \text{constant} + e \log(1 - MTR_{1,t}) + b_1 \text{trend} + b_2 \text{trend}^2 + b_3 \text{trend}^3 + \varepsilon_t$$

Models	$a$	$e$	$b_1$	$b_2$	$b_3$	$R^2$	$D - W$	$S.E.E.$
Top 95-99 %								
No time trend	-0.918 (-102.5)	0.061 (2.42)				0.168	0.453	0.009
Linear time trend	-0.946 (-25.6)	0.003 (-0.04)	0.000 (0.78)			0.186	0.457	0.010
Linear square trend	-0.901 (-25.3)	0.073 (0.99)	-0.002 (-2.14)	0.000 (3.03)		0.392	0.625	0.008
Cube time trend	-0.939 (-41.7)	-0.050 (-1.05)	-0.008 (-7.60)	0.001 (7.51)	0.000 (-6.80)	0.748	1.460	0.005

$$\log(S_{k,t}) = \text{constant} + e \log(1 - MTR_{k,t}) + b_1 \text{trend} + b_2 \text{trend}^2 + b_3 \text{trend}^3 + \varepsilon_t$$

Models	$a$	$e$	$b_1$	$b_2$	$b_3$	$R^2$	$D - W$	$S.E.E.$
Top 1 % as control top 95-99 %								
No time trend	0.000 (0.09)	1.131 (5.43)				0.513	1.941	0.026
Linear time trend	0.001 (0.07)	1.311 (5.23)	0.000 (-0.02)			0.513	1.942	0.027
Linear square trend	-0.006 (-0.39)	1.218 (4.99)	0.001 (0.54)	0.000 (-0.56)		0.518	1.978	0.027
Cube time trend	0.007 (0.32)	1.309 (5.02)	-0.004 (-0.56)	0.000 (0.72)	0.000 (-0.82)	0.531	2.040	0.027

$$[\log(\frac{S_{1,t}}{S_{k,t}}) - \log(\frac{S_{1,t-1}}{S_{k,t-1}})] = \text{constant} + e [\log(\frac{1-MTR_{1,t}}{1-MTR_{k,t}}) - \log(\frac{1-MTR_{1,t-1}}{1-MTR_{k,t-1}})] + b_1 \text{trend} + b_2 \text{trend}^2 + b_3 \text{trend}^3 + \varepsilon_t$$

Parameter estimates:  $a$  = constant,  $e$  = elasticity,  $b_1$  = linear time trend,  $b_2$  = square time trend,  $b_3$  = cube time trend.